Package 'MAGEE'

November 18, 2021

Type Package

Version 1.0.1

Date 2021-11-13

Title Mixed Model Association Test for GEne-Environment Interaction

Description Use a 'glmmkin' class object (GMMAT package) from the null model to perform generalized linear mixed model-based single-variant and variant set main effect tests, gene-environment interaction tests, and joint tests for association, as proposed in Wang et al. (2020) <DOI:10.1002/gepi.22351>.

License GPL-3

Copyright See COPYRIGHTS for details.

Imports Rcpp, Matrix, parallel, MASS, SeqArray, SeqVarTools, foreach, GMMAT, CompQuadForm

Suggests doMC, testthat

LinkingTo Rcpp, RcppArmadillo

Encoding UTF-8

LazyData true

Depends R (>= 3.2.0)

NeedsCompilation yes

Author Xinyu Wang [aut],
Han Chen [aut, cre],
Duy Pham [aut]

Maintainer Han Chen <han.chen.2@uth.tmc.edu>

Repository CRAN

Date/Publication 2021-11-18 07:40:05 UTC

R topics documented:

`MAGEE-package` *Mixed Model Association Test for GEne-Environment Interaction*

Description

An R package for performing generalized linear mixed model-based single-variant and variant set main effect tests, gene-environment interaction tests, and joint tests for association, as proposed in Wang et al. (2020) <DOI:10.1002/gepi.22351>.

Details

Package:	MAGEE
Type:	Package
Version:	1.0.1
Date:	2021-11-13
License:	GPL-3

Author(s)

Xinyu Wang, Han Chen, Duy T. Pham

Maintainer: Han Chen <Han.Chen.2@uth.tmc.edu>

References

Wang, X., Lim, E., Liu, C, Sung, Y.J., Rao, D.C., Morrison, A.C., Boerwinkle, E., Manning, A. K., and Chen, H. (2020) Efficient gene-environment interaction tests for large biobank-scale sequencing studies. Genetic Epidemiology, 44(8): 908-923.

Chen, H., Huffman, J.E., Brody, J.A., Wang, C., Lee, S., Li, Z., Gogarten, S.M., Sofer, T., Bielak, L.F., Bis, J.C., et al. (2019) Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies. The American Journal of Human Genetics, 104 (2): 260-274.

`example` *Example dataset*

Description

Example dataset for MAGEE.

Format

Contains the following objects:

pheno a data frame of 400 observations from a cross-sectional study with 5 variables: id, disease, trait, age and sex.

pheno2 a data frame of 2,000 observations from a longitudinal study with 400 individuals and 5 variables: id, y.repeated, y.trend, time and sex.

GRM a genetic relationship matrix for 400 observations.

`glmm.gei` *GLMM based single variant tests for gene-environment interactions*

Description

Use a glmmkin class object from the null GLMM to perform single variant main effect score test, gene-environment interaction test, or joint test for association with genotypes in a GDS file .gds.

Usage

```
glmm.gei(null.obj, interaction, geno.file, outfile, bgen.samplefile = NULL,
  interaction.covariates = NULL,  meta.output = F, center=T, MAF.range = c(1e-7, 0.5),
   miss.cutoff = 1, missing.method = "impute2mean", nperbatch=100, ncores = 1,
   verbose = FALSE)
```

Arguments

`null.obj`	a class glmmkin object, returned by fitting the null GLMM using `glmmkin( )`.
`interaction`	a numeric or a character vector indicating the environmental factors. If a numeric vector, it represents which indices in the order of covariates are the environmental factors; if a character vector, it represents the variable names of the environmental factors.
`geno.file`	the full name of a GDS file (including the suffix .gds).
`outfile`	the output file name.
`bgen.samplefile`	path to the BGEN .sample file. Required when the BGEN file does not contain sample identifiers.
`interaction.covariates`	a numeric or a character vector indicating the interaction covariates. If a numeric vector, it represents which indices in the order of covariates are the interaction covariates; if a character vector, it represents the variable names of the interaction covariates.
`meta.output`	boolean value to modiy the output file.If TRUE, the GxE effect estimate and variance and covariance associated with the effect estimate are included in the output file. (default = FALSE)

center	a logical switch for centering genotypes before tests. If TRUE, genotypes will be centered to have mean 0 before tests, otherwise raw values will be directly used in tests (default = TRUE).
MAF.range	a numeric vector of length 2 defining the minimum and maximum minor allele frequencies of variants that should be included in the analysis (default = c(1e-7, 0.5)).
miss.cutoff	the maximum missing rate allowed for a variant to be included (default = 1, including all variants).
missing.method	method of handling missing genotypes.Either "impute2mean" or "omit" (default = "impute2mean").
nperbatch	an integer for how many SNPs should be tested in a batch (default = 100). The computational time can increase dramatically if this value is either small or large. The optimal value for best performance depends on the user's system.
ncores	a positive integer indicating the number of cores to be used in parallel computing (default = 1).
verbose	a logical switch for whether failed matrix inversions should be written to outfile.err for debugging (default = FALSE).

Value

NULL

Author(s)

Xinyu Wang, Han Chen, Duy Pham

References

Chen, H., Wang, C., Conomos, M.P., Stilp, A.M., Li, Z., Sofer, T., Szpiro, A.A., Chen, W., Brehm, J.M., Celedón, J.C., Redline, S., Papanicolaou, G.J., Thornton, T.A., Laurie, C.C., Rice, K. and Lin, X. (2016) Control forpopulation structure and relatedness for binary traits in genetic association studies via logistic mixed models. The American Journal of Human Genetics 98, 653-666.

Examples

```
library(GMMAT)
data(example)
attach(example)

model0 <- glmmkin(disease ~ age + sex, data = pheno, kins = GRM,
                  id = "id", family = binomial(link = "logit"))

infile <- system.file("extdata", "geno.gds", package = "MAGEE")
gds_outfile <- tempfile()
glmm.gei(model0, interaction='sex', geno.file = infile, outfile = gds_outfile)

infile <- system.file("extdata", "geno.bgen", package = "MAGEE")
samplefile <- system.file("extdata", "geno.sample", package = "MAGEE")
bgen_outfile <- tempfile()
```

```
glmm.gei(model0, interaction='sex', geno.file = infile, outfile = bgen_outfile,
         bgen.samplefile = samplefile)

unlink(paste0("glmm.gei.", c("gds","bgen"), bgen_outfile, gds_outfile))
```

MAGEE *Mixed model Association tests for GEne-Environment interactions*

Description

Use a glmmkin class object from the null GLMM to perform variant set-based main effect tests, gene-environment interaction tests, and joint tests for association with genotypes in a GDS file (.gds). 7 user-defined tests are included: Main effect variance component test (MV), Main effect hybrid test of burden and variance component test using Fisher's method (MF), Interaction variance component test (IV), Interaction hybrid test of burden and variance component test using Fisher's method (IF), Joint variance component test (JV), Joint hybrid test of burden and variance component test using Fisher's method (JF), and Joint hybrid test of burden and variance component test using double Fisher's procedures (JD).

Usage

```
MAGEE(null.obj, interaction, geno.file, group.file,  group.file.sep = "\t",
    bgen.samplefile = NULL, interaction.covariates = NULL, meta.file.prefix = NULL,
      MAF.range = c(1e-7, 0.5), MAF.weights.beta = c(1, 25), miss.cutoff = 1,
      missing.method = "impute2mean", method = "davies", tests = "JF",
      use.minor.allele = FALSE, auto.flip = FALSE,
      Garbage.Collection = FALSE, is.dosage = FALSE, ncores = 1)

MAGEE.prep(null.obj, interaction, geno.file, group.file, interaction.covariates = NULL,
           group.file.sep = "\t", auto.flip = FALSE)

MAGEE.lowmem(MAGEE.prep.obj, geno.file = NULL, meta.file.prefix = NULL,
          MAF.range = c(1e-7, 0.5), MAF.weights.beta = c(1, 25), miss.cutoff = 1,
             missing.method = "impute2mean", method = "davies", tests = "JF",
         use.minor.allele = FALSE, Garbage.Collection = FALSE, is.dosage = FALSE,
     ncores = 1)
```

Arguments

null.obj	a class glmmkin object, returned by fitting the null GLMM using glmmkin( ).
interaction	a numeric or a character vector indicating the environmental factors. If a numberic vector, it represents which indices in the order of covariates are the environmental factors; if a character vector, it represents the variable names of the environmental factors.
geno.file	a .gds file for the full genotypes. The sample.id in geno.file should overlap id_include in null.obj. It is recommended that sample.id in geno.file include the full samples (at least all samples as specified in id_include of

null.obj). It is not necessary for the user to take a subset of geno.file before running the analysis.

group.file a plain text file with 6 columns defining the test units. There should be no headers in the file, and the columns are group name, chromosome, position, reference allele, alternative allele and weight, respectively.

group.file.sep the delimiter in group.file (default = "\t").

bgen.samplefile
path to the BGEN .sample file. Required when the BGEN file does not contain sample identifiers.

interaction.covariates
a numeric or a character vector indicating the interaction covariates. If a numeric vector, it represents which indices in the order of covariates are the interaction covariates; if a character vector, it represents the variable names of the interaction covariates.

meta.file.prefix
the prefix for meta-analysis (default = "NULL").

MAF.range a numeric vector of length 2 defining the minimum and maximum minor allele frequencies of variants that should be included in the analysis (default = c(1e-7, 0.5)).

MAF.weights.beta
a numeric vector of length 2 defining the beta probability density function parameters on the minor allele frequencies. This internal minor allele frequency weight is multiplied by the external weight given by the group.file. To turn off internal minor allele frequency weight and only use the external weight given by the group.file, use c(1, 1) to assign flat weights (default = c(1, 25)).

miss.cutoff the maximum missing rate allowed for a variant to be included (default = 1, including all variants).

missing.method method of handling missing genotypes. Either "impute2mean" or "impute2zero" (default = "impute2mean").

method a method to compute p-values for the test statistics (default = "davies"). "davies" represents an exact method that computes a p-value by inverting the characteristic function of the mixture chisq distribution, with an accuracy of 1e-6. When "davies" p-value is less than 1e-5, it defaults to method "kuonen". "kuonen" represents a saddlepoint approximation method that computes the tail probabilities of the mixture chisq distribution. When "kuonen" fails to compute a p-value, it defaults to method "liu". "liu" is a moment-matching approximation method for the mixture chisq distribution.

tests a character vector indicating which MAGEE tests should be performed ("MV" for the main effect variance component test, "MF" for the main effect combined test of the burden and variance component tests using Fisher's method, "IV" for the interaction variance component test, "IF" for the interaction combined test of the burden and variance component tests using Fisher's method, "JV" for the joint variance component test for main effect and interaction, "JF" for the joint combined test of the burden and variance component tests for main effect and interaction using Fisher's method, or "JD" for the joint combined test of the burden and variance component tests for main effect and interaction using

double Fisher's method.). The "MV" and "IV" test are automatically included when performing "JV", and the "MF" and "IF" test are automatically included when performing "JF" or "JD" (default = "JF").

`use.minor.allele`
: a logical switch for whether to use the minor allele (instead of the alt allele) as the coding allele (default = FALSE). It does not change "MV", "IV", and "JV" results, but "MF", "IF", "JF", and "JD" results will be affected.Along with the MAF filter, this option is useful for combining rare mutations, assuming rare allele effects are in the same direction.

`auto.flip`
: a logical switch for whether to enable automatic allele flipping if a variant with alleles ref/alt is not found at a position, but a variant at the same position with alleles alt/ref is found (default = FALSE). Use with caution for whole genome sequence data, as both ref/alt and alt/ref variants at the same position are not uncommon, and they are likely two different variants, rather than allele flipping.

`Garbage.Collection`
: a logical switch for whether to enable garbage collection in each test (default = FALSE). Pay for memory efficiency with slower computation speed.

`is.dosage`
: a logical switch (default = FALSE).

`ncores`
: a positive integer indicating the number of cores to be used in parallel computing (default = 1).

`MAGEE.prep.obj`
: a class MAGEE.prep object, returned by `MAGEE.prep`.

Value

a data frame with the following components:

`group`
: name of the test unit group.

`n.variants`
: number of variants in the test unit group that pass the missing rate and allele frequency filters.

`miss.min`
: minimum missing rate for variants in the test unit group.

`miss.mean`
: mean missing rate for variants in the test unit group.

`miss.max`
: maximum missing rate for variants in the test unit group.

`freq.min`
: minimum coding allele frequency for variants in the test unit group.

`freq.mean`
: mean coding allele frequency for variants in the test unit group.

`freq.max`
: maximum coding allele frequency for variants in the test unit group.

`freq.strata.min`
: minimum coding allele frequency of each strata if the environmental factor is categorical.

`freq.strata.max`
: maximum coding allele frequency of each strata if the environmental factor is categorical.

`MV.pval`
: MV test p-value.

`MF.pval`
: MF test p-value.

`IV.pval`
: IV test p-value.

IF.pval	IF test p-value.
JV.pval	JV test p-value.
JF.pval	JF test p-value.
JD.pval	JD test p-value.

Author(s)

Xinyu Wang, Han Chen, Duy Pham

References

Wang, X., Lim, E., Liu, C, Sung, Y.J., Rao, D.C., Morrison, A.C., Boerwinkle, E., Manning, A. K., and Chen, H. (2020) Efficient gene-environment interaction tests for large biobank-scale sequencing studies. Genetic Epidemiology, 44(8): 908-923. Chen, H., Huffman, J.E., Brody, J.A., Wang, C., Lee, S., Li, Z., Gogarten, S.M., Sofer, T., Bielak, L.F., Bis, J.C., et al. (2019) Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies. The American Journal of Human Genetics, 104 (2): 260-274.

Examples

```
library(GMMAT)
    data(example)
    attach(example)
    model0 <- glmmkin(disease ~ age + sex, data = pheno, kins = GRM,
    id = "id", family = binomial(link = "logit"))
    geno.file <- system.file("extdata", "geno.gds", package = "MAGEE")
    group.file <- system.file("extdata", "SetID.withweights.txt", package = "MAGEE")
    out <- MAGEE(model0, interaction='sex', geno.file, group.file, group.file.sep = "\t",
                 tests=c("JV", "JF", "JD"))
    print(out)
```

Index